

## References on quantization

- [1] S. Jayant and P. Noll, *Digital Coding of Waveforms*, New Jersey, Prentice Hall, Chapter 5, pp. 221–251, 1984.
- [2] A. Gersho, “Quantization,” *IEEE Communications Society Magazine*, pp. 20–29, September, 1977.
- [3] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Pearson, 4th edition, 2018.

## References on the simplest uniform quantization

- [1] S. Na, D. Neuhoff, “Monotonicity of step sizes of MSE-optimal symmetric uniform scalar quantizers,” *IEEE Transaction on Information Theory*, vol. 65, no. 3, pp. 1782–1792, 2019.
- [2] S. Na, D. Neuhoff, “On the convexity of the MSE distortion of symmetric uniform scalar quantization,” *IEEE Transaction on Information Theory*, vol. 64, no.4, pp. 2626–2638, 2018.
- [3] D. Hui, D. L. Neuhoff, “Asymptotic analysis of optimal fixed-rate uniform scalar quantization,” *IEEE Transaction on Information Theory*, vol. 47, no. 3, pp. 957–977, 2001.

## References on companding quantization

- [1] Z. Perić, J. Nikolić, “High-quality Laplacian source quantisation using a combination of restricted and unrestricted logarithmic quantisers,” *IET Signal Processing*, vol. 6, no. 7, pp. 633–640, 2012.
- [2] Z. Perić, M. Petković, J. Nikolić, “Optimization of multiple region quantizer for Laplacian source,” *Digital Signal Processing*, vol. 27, pp. 150–158, 2014.
- [3] J. Nikolić, Z. Perić, A. Jovanović, “Two forward adaptive dual-mode companding scalar quantizers for Gaussian source,” *Signal Processing*, vol. 120, pp. 129–140, 2016.
- [4] D. Aleksić, Z. Perić, “Analysis and design of robust quasilogarithmic quantizer for the purpose of traffic optimisation,” *Information Technology and Control*, vol. 47, no.4, pp. 615–622, 2018.
- [5] Z. Perić, D. Aleksić, “Quasilogarithmic quantizer for Laplacian source: support region ubiquitous optimization task,” *Revue Roumaine des Sciences Techniques*, vol. 64, no. 4, pp. 403–408, 2019.
- [6] S. Na, D. Neuhoff, “On the support of MSE-optimal, fixed-rate, scalar quantizers,” *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2972–2982, 2001.
- [7] S. Na, “Asymptotic formulas for mismatched fixed-rate minimum MSE Laplacian quantizers,” *IEEE Signal Processing Letters*, vol. 15, pp. 13–16, 2008.
- [8] S. Na, “On the support of fixed-rate minimum mean-squared error scalar quantizers for a Laplacian source,” *IEEE Transactions on Information Theory*, vol. 50, No. 5, pp. 937–944, May, 2004.

## References on neural networks and quantization of neural networks

- [1] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- [2] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2007.
- [3] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, Springer 2014.
- [4] S. Raschka, V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*, Packt Publishing, 3rd Edition, 2019.
- [5] U. Spagnolini, *Statistical Signal Processing in Engineering*, Wiley, 2018.

- [6] V. Delić, Z. Perić, M. Sečujski, N. Jakovljević, J. Nikolić, et al., “Speech technology progress based on new machine learning paradigm,” *Computational Intelligence and Neuroscience*, Article ID 4368036, 19 pages, 2019.
- [7] X. Long, X. R. Zeng, Z. Ben et al., “A novel low-bit quantization strategy for compressing deep neural networks,” *Computational Intelligence and Neuroscience*, Article ID 7839064, 2020.
- [8] R. Banner, Y. Nahshan, D. Soudry, “Postraining 4-bit quantization of convolutional networks for rapid-deployment,” *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, Canada, pp. 7948–7956, December 2019.
- [9] N. Nicodemo, G. Naithani, K. Drossos, T. Virtanen, R. Saletti, “Memory requirement reduction of deep neural networks for field programmable gate arrays using low-bit quantization of parameters,” *28th European Signal Processing Conference EUSIPCO*, Amsterdam, pp. 466–470, January 2021.
- [10] I. Hubara, M. Courbariaux, D. Soudry, et al., “Binarized neural networks,” *30th Conference on Neural Information Processing Systems*, Barcelona, Spain, 2016, pp. 1–9. December 2016.
- [11] D. Lin, S. Talathi, S. Annapureddy, “Fixed point quantization of deep convolutional networks,” *33rd International Conference on Machine Learning*, New York, NY, USA, 2016, 48, pp. 2849–2858, June 2016.
- [12] I. Hubara, M. Courbariaux, D. Soudry, et al., “Quantized neural networks: Training neural networks with low precision weights and activations,” *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6869–6898, 2017.
- [13] K. Huang, B. Ni, X. Yang, “Efficient quantization for neural networks with binary weights and low bit width activations,” *Thirty-Third AAAI Conference on Artificial Intelligence*, Honolulu, HI, USA, 2019, 33, (01), pp. 3854–3861. 27 January–1 February 2019.
- [14] J. Yang, X. Shen, J. Xing, et al., “Quantization networks,” *IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019, pp. 7308–7316. June 2019.
- [15] W. Zha, G. Venayagamoorthy, “Neural networks based non-uniform scalar quantizer design with particle swarm optimization,” *2005 IEEE Swarm Intelligence Symposium*, Pasadena, CA, USA, 2005, pp. 143–148. 2005.
- [16] S. Uhlich, L. Mauch, F. Cardinaux, et al., “Mixed precision DNNs: All you need is a good parametrization,” *8th International Conference on Learning Representations*, Addis Ababa, Ethiopia, 2020, pp. 1–20. April 2020.
- [17] S. Gupta, S. Ullah, K. Ahuja, A. Tiwari, A. Kumar, “ALigN: A highly accurate adaptive layerwise  $\log_2$  lead quantization of pre-trained neural networks,” *IEEE Access*, vol. 8, pp. 118899–118911, 2020.
- [18] S. Sanghyun, K. Juntae, “Efficient weights quantization of convolutional neural networks using kernel density estimation based non-uniform quantizer,” *Applied Sciences*, vol. 9, pp. 2559, 2019.
- [19] N. Naz, A. H. Malik, A. B. Khurshid et al., “Efficient processing of image processing applications on CPU/GPU,” *Mathematical Problems in Engineering*, vol. 2020, Article ID 4839876, 14 pages, 2020.

## References on EM algorithm and clustering

- [1] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006, Chapter 9.
- [2] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, 2016.
- [3] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Morgan Kaufmann Series in Data Management Systems, 2016.

- [4] A. Cord, C. Ambroise, J. P. Cocquerez, “Feature selection in robust clustering based on Laplace mixture,” *Pattern Recognition Letters*, vol. 27, no. 6, pp. 627–635, 2006.
- [5] A. P. Dempster, N. M. Laird, D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.
- [6] S. Gamini, Y. Veena, “Analysis of k-means clustering on uniform and non-uniform data set,” *International Journal of Advances in Electronics and Computer Science*, vol. 3, no. 12, pp. 2393–2835, 2016.
- [7] X. Lin, X. Yang, Y. Li, “A deep clustering algorithm based on Gaussian mixture model,” *ISAI 2019, Journal of Physics: Conference Series*, 1302, 2019.
- [8] M. S. Yang, C. Y. Lai, C. Y. Lin, “A robust EM clustering algorithm for Gaussian mixture models,” *Pattern Recognition*, vol. 45, no. 11, pp. 3950–3961, 2012.
- [9] A. Cedric, L. A. John, V. Michael, “On convergence problems of the EM algorithm for finite Gaussian mixtures,” *ESANN 2003, 11th European Symposium on Artificial Neural Networks*, Bruges, Belgium, 2003, pp. 99–106.
- [10] I. H. Tseng, O. Verscheure, D. S. Turaga, U. V. Chaudhari, “Quantization for adapted GMM-based speaker verification,” *IEEE ICASSP*, Toulouse, France, May 2006.
- [11] I. H. Tseng, O. Verscheure, D. S. Turaga, U. V. Chaudhari, “Optimized one-bit quantization for adapted GMM-based speaker verification,” *INTERSPEECH*, Antwerp, Belgium, August 27–31, 2007, pp. 786–789.

## References on variable length coding

- [1] Z. Perić, J. Nikolić, D. Milić, “Variable-length coding for performance improvement of asymptotically optimal unrestricted polar quantization of bivariate Gaussian source,” *Information Sciences*, vol. 252, no. 10, pp. 157–167, 2013.
- [2] Z. Perić, J. Nikolić, A. Mosić, M. Petković, “Design of fixed and adaptive companding quantizer with variable-length codeword for memoryless Gaussian source,” *Informatica*, vol. 24, no. 1, pp. 71–86, 2013.
- [3] Z. Perić, J. Nikolić, L. Velimirović, M. Stanković, D. Aleksić, “Asymmetrical two-level scalar quantizer with extended Huffman coding for compression of Laplacian source,” *Advances in Electrical and Computer Engineering*, vol. 13, no. 2, pp. 39–42, 2013.
- [4] Z. Perić, J. Nikolić, A. Mosić, “Design of forward adaptive hybrid quantiser with Golomb-Rice code for compression of Gaussian source,” *IET Communications*, vol. 8, no. 3, pp. 372–377, 2014.

## References on application of DNN in the audio domain

- [1] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S. Chang, and T. Sainath, “Deep learning for audio signal processing,” *J. Selected Topics Sig. Proc.*, vol. 13, no. 2, pp. 206–219, May 2019.
- [2] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (Adaptive Computation and Machine Learning series), MIT Press, 2016.
- [3] J. Abeßer, “A review of deep learning based methods for acoustic scene classification,” *Appl. Sciences*, vol. 10, 16 pages, 2020.
- [4] C. Roletscheck, T. Watzka, A. Seiderer, D. Schiller, and E. André, “Using an evolutionary approach to explore convolutional neural networks for acoustic scene classification,” *Detection Class. Acoust. Scenes Events Workshop (DCASE)*, New York, NY, USA, 25–26 Oct. 2019.

- [5] F. Roche, T. Hueber, S. Linier and L. Girin, "Autoencoders for music sound modeling: a comparison of linear, shallow, deep, recurrent and variational models," arXiv:1806.04096, Jun. 2018.
- [6] M. Espi, M. Fujimoto, K. Kinoshita, and T. Nakatani, "Exploiting spectro-temporal locality in deep learning based acoustic event detection," *EURASIP J. Aud., Speech, Music Proc.*, article no. 26, 2015.
- [7] I. McLoughlin, Z. Xie, Y. Song, H. Phan, and R. Palaniappan, "Time–frequency feature fusion for noise robust audio event classification," *Circuits, Sys., Sig. Proc.*, vol. 39, pp. 1672–1687, 2020.
- [8] I. W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–25, Apr. 2017.
- [9] A. Khamparia, D. Gupta, N. G. Nguyen, A. Khanna, B. Pandey, and P. Tiwari, "Sound classification using convolutional neural network and tensor deep stacking network," *IEEE Access*, vol. 7, pp. 7717 – 7727, Jan. 2019.
- [10] J. Pons, O. Slizovskaia, R. Gong, E. Gomez, and X. Serra, "Timbre analysis of music audio signals with convolutional neural networks," arXiv:1703.06697, 2017.
- [11] O. Abdel-Hamid, A. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Trans. Audio Speech Lang. Proc.*, vol. 22, no. 10, pp. 1533–1545, 2014.
- [12] Y. Zeng, H. Mao, D. Peng, and Z. Yi, "Spectrogram based multi-task audio classification," *Multimedia Tools App.*, vol. 78, no. 3, pp. 3705–3722, 2019.
- [13] S. Hershey, et al. "CNN architectures for large-scale audio classification," *IEEE Int. Conf. Acoust. Speech Sig. Proc. (ICASSP)*, New Orleans, USA, 5–9 Mar. 2017.
- [14] P. Verma, J. O. Smith, "Neural style transfer for audio spectrograms," *31st Conf. Neural Inf. Proc. Systems (NIPS 2017)*, Long Beach, CA, USA, Dec. 2017.
- [15] J. Lee, J. Park, K. L. Kim, and J. Nam, "Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms," *Sound Music Comput. Conf.*, Espoo, Finland, 5–8 Jul. 2017.
- [16] S. Durand, J. P. Bello, B. David, and G. Richard, "Robust downbeat tracking using an ensemble of convolutional networks," *IEEE ACM Trans. Audio Speech Lang. Proc.*, vol. 25, no. 1, pp. 72–85, 2017.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *IEEE Conf. Comput. Vision Pattern Recognition*, Las Vegas, USA, pp. 770–778, 27-30 Jun. 2016.
- [18] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [19] S. Mun, S. Shon, W. Kim, D. K. Han, and H. Ko, "Deep neural network based learning and transferring mid-level audio features for acoustic scene classification," *IEEE Int. Conf. Acoust. Speech Sig. Proc. (ICASSP)*, New Orleans, LA, USA, pp. 796–800, 5–9 Mar. 2017.
- [20] S. H. Bae, I. Choi, and N. S. Kim, "Acoustic scene classification using parallel combination of LSTM and CNN," *Detection Class. Acoust. Scenes Events Workshop (DCASE)*, Budapest, Hungary, 3 Sep. 2016.
- [21] I. Goodfellow, et al. "Generative adversarial nets," *27th Int. Conf. Neural Information Proc. Systems (NIPS)*, Montreal, Canada, vol. 2., pp. 2672–2680, 8-13 Dec. 2014.

- [22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [23] J. Schlüter and S. Böck, "Improved musical onset detection with convolutional neural networks," *IEEE Int. Conf. Acoust. Speech Sig. Proc. (ICASSP)*, Florence, Italy, pp. 6979–6983, 4-9 May 2014.
- [24] Z. Ouyang, H. Yu, W-P. Zhu, and B. Champagne, "A fully convolutional neural network for complex spectrogram processing in speech enhancement," *IEEE Int. Conf. Acoust. Speech Sig. Proc. (ICASSP)*, Brighton, UK, 12–17 May 2019.
- [25] Y-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," *27th Int. Conf. Machine Learning*, Haifa, Israel, pp. 111–118, 21-24 Jun. 2010.
- [26] J. Sharma, O. C. Granmo, and M. Goodwin, "Environment sound classification using multiple feature channels and deep convolutional neural networks," *Journal Latex Class Files*, vol. 14, no. 8, 11 pages, 2015.
- [27] J. Cho, S. Yun, H. Park, J. Eum, and K. Hwang, "Acoustic scene classification based on a large-margin factorized CNN," *Detection Class. Acoust. Scenes Events Workshop (DCASE)*, New York, NY, USA, pp. 45–49, 25–26 Oct. 2019.
- [28] A. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," in *SSW*, vol. 125, (arXiv:1609.03499), 2016.
- [29] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, *Wavelets, Time-Frequency Methods and Phase Space*, Springer, pp. 289–297, 1989.
- [30] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Tran. Pattern Analysis Machine Intell.*, vol. 40, pp. 834-848, Apr. 2018.
- [31] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *Int. Conf. Learning Representations (ICLR)*, San Juan, Puerto Rico, 2–4 May 2016.
- [32] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98-113, Jan. 1997.
- [33] J. Naranjo-Alcazar, S. Perez-Castanos, P. Zuccarello, F. Antonacci, and M. Cobos, "Open set audio classification using autoencoders trained on few data," *Sensors*, vol. 20, 3741 (19 pages), 2020.
- [34] M. Tagliasacchi, B. Gfeller, F.d.C. Quitry, D. Roblek, "Self-supervised audio representation learning for mobile devices," arXiv:1905.11796, 2019.
- [35] F.d.C. Quitry, M. Tagliasacchi, D. Roblek, "Learning audio representations via phase prediction," arXiv:1910.11910, 2019.